Celal Bayar University Journal of Science

The Reliability of Citizen Science Data: Testing A Data Set from eBird Türkiye with Benford's Law

Çetin Çelik[1,2] (iD), Murat Afsar[3]* (iD)

[1] PhD. Student, Manisa Celal Bayar University Graduate School, Yunusemre-Manisa-Türkiye
[2] Asistant Manager, Akhisar Bilim ve Sanat Merkezi, Akhisar-Manisa-Türkiye
[3] Manisa Celal Bayar Üniversitesi, Faculty of Engineering and Natural Sciences, Biology Department, Zooloji Section
* murat.afsar@cbu.edu.tr
* Orcid No: 0000-0003-3210-8262

Abstract

Citizen science is a rapidly evolving and increasingly popular field of scientific inquiry in which data are collected by volunteers (citizens). Applications of citizen science represent a growing data resource used by researchers, nature conservationists, and public institutions. Although citizen science data can make substantial contributions within the scope of Conservation Biology—particularly to the protection and management of biological diversity, ecosystems, and natural resources—concerns regarding data quality persist, despite the expanding potential of citizen science. Ensuring that data collected through citizen science efforts are accurate and reliable is crucial for conducting valid assessments in Conservation Biology and for enabling more effective conservation practices.In this study, we sought to test the accuracy of bird observation data gathered via the citizen science platform eBird Türkiye (http://www.ebird.org)—which contains observation counts for 533 bird species recorded in Türkiye—by applying Benford's Law. Benford's Law is a mathematical approach developed to analyze the distribution of numbers in data sets that may not have a strictly "natural" origin. Additionally, a Chi-Square Goodness-of-Fit Test was conducted to detect differences between the observed (actual) values and the expected (Benford's Law) values. The results indicate that the tested data from the eBird Türkiye platform conform to Benford's Law.

Keywords: Benford Law, Conservation Biology, eBird, Citizen Science, Biodiversity

1. Introduction

The rapid pace of scientific advancement has led to increasingly specialized fields of expertise. Although specialization supports scientific progress, it also fosters a structure in which only experts produce knowledge, making it challenging for the broader public to grasp new developments. One response to this issue—particularly prominent in Europe—is the institutionalization and encouragement of public participation in scientific research under the banner of "Citizen Science," which aligns with open science principles. In addition to reforming scientific methodologies, this approach seeks to enhance public engagement in scientific processes [1].

Citizen science is a method that promotes participation by non-professional individuals in scientific research. By involving volunteers, it reduces research costs, facilitates faster project completion, strengthens communication between science and society, and increases scientific literacy. It also encourages a more collaborative form of research by contributing to participants' self-realization processes [1]. Owing to its potential for advancing scientific literacy and involving diverse segments of society in research, citizen science has grown increasingly popular in recent years [2].

Citizen science applications have emerged as a rapidly expanding data resource, widely used by researchers, nature conservationists, and public institutions [3, 4, 5, 6, 7]. The data collected by volunteers are vital, given the increased taxonomic, geographical, and temporal scope [8].

However, one of the greatest challenges in citizen science lies in ensuring data quality and the accuracy of

volunteer-collected information [2]. Despite the growing potential of citizen science, concerns regarding data reliability remain [9]. To use these data effectively in research, conservation, and policy-making, their reliability must be carefully evaluated. Data quality issues often reflect spatial, temporal, and taxonomic biases [10] or insufficient sampling [11, 12, 13]. Consequently, developing and applying methods to enhance the reliability of citizen science data is essential.

A recent Web of Science categorization of publications using the keyword "Citizen Science" reveals that Environmental Sciences has the highest number of related publications [1] as shown in Figure 1.
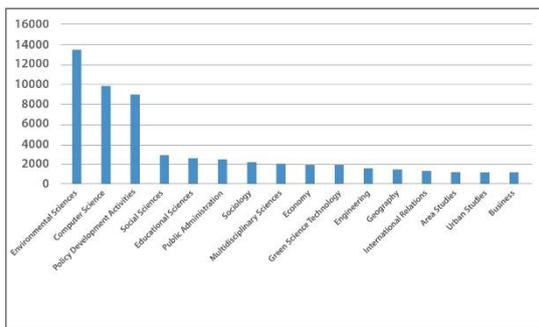

Figure 1. Citizen Science Publications by Scientific Field [1] (Reproduced from the original source).

To understand why citizen science is so significant to environmental sciences and the advantages it offers, one must first examine the concept of "Conservation Biology." Conservation Biology is an interdisciplinary field focused on the protection and sustainable management of biological diversity, ecosystems, and natural resources. It integrates ecological, evolutionary, and socio-economic factors through various approaches to address complex issues involving the preservation of natural systems. One of its core aims is to prevent the loss of biological diversity by protecting and restoring threatened or endangered species and habitats. Simultaneously, Conservation Biology deals with the root causes of biodiversity decline, such as habitat destruction, pollution, climate change, and overexploitation of resources. This endeavor requires collaboration among scientists from different disciplines, policy makers, and other stakeholders. Conservation biologists investigate how human activities affect natural systems in order to develop effective, sustainable strategies for resource management [14].

In this context, citizen science offers critical support to conservation biology. The large-scale data collected by volunteers serve as an essential resource for monitoring biodiversity, recognizing threats, and shaping conservation strategies. Beyond aiding scientific research, citizen science also boosts public engagement in nature conservation by inviting active community participation, thereby increasing environmental awareness and promoting scientific literacy. Consequently, citizen science not only democratizes the production of scientific knowledge but also broadens the social base for nature conservation initiatives, generating benefits for both science and society.

Numerous statistical methods have been used to evaluate the accuracy and reliability of citizen science data. In this study, we applied Benford's Law, a mathematical technique designed to analyze the distribution of numbers in data sets potentially lacking a natural origin [15, 16]. This method tests whether a given data set follows the leading digit distribution predicted by Benford's Law. Because Benford's Law is a powerful tool for assessing the trustworthiness of numerical data and detecting inconsistencies, it has been applied in multiple fields, including finance and the natural sciences.

Benford's Law has proven effective in several sectors for checking the reliability of reported natural resource harvesting data (for instance, in fishing or tropical hunting; 17) and for detecting illegal deforestation activities [18]. This approach is also employed to study naturally occurring biological data sets: prior research has demonstrated that certain biological data—such as cyanobacterial cell counts [19] pollen counts [20], genome sizes [21] and angiosperm taxon counts [22]—can align with Benford's Law. Overall, Benford's Law and similar digit-based tests have been recommended as simple yet effective preliminary screening methods for large, complex ecological data sets [20, 23, 24]. Given the increasing complexity and volume of ecological data—including citizen science data—Benford's Law has significant potential for identifying anomalies or errors in naturally occurring data sets, thus providing critical insights for ecological and environmental research [25]. In particular, this method is valuable for evaluating the reliability of large-scale data sets and for deepening our understanding of ecological processes.

Besides Benford's Law, alternative methods for assessing the accuracy of citizen science data include "Inter-Observer Agreement," "Expert Validation," and "Outlier Detection." Inter-Observer Agreement compares data collected by various volunteers to measure consistency, often using statistical measures such as Cohen's Kappa or Fleiss's Kappa [26]. Expert Validation involves detailed reviews by specialists, who compare the volunteer-collected data with those gathered through conventional scientific techniques [26]. Meanwhile, Outlier Detection identifies data points that differ substantially from the rest of the dataset, helping pinpoint—and potentially eliminate—errors [26].

When compared with other European and Middle Eastern countries, Türkiye's remarkable biological diversity in relation to its size positions it as one of the most

biodiverse nations in the region. Thus, Türkiye has a major need for citizen science initiatives aimed at identifying and conserving this biodiversity. Nonetheless, citizen science remains only partly developed here. Platforms such as Trakuş, AdaMerOs, and eBird Türkiye represent major strides in this domain by harnessing volunteer contributions to gather and monitor biodiversity information, highlighting the potential impact of citizen science in Türkiye. Given the country's notable biodiversity, expanding and developing citizen science activities could significantly bolster its conservation efforts.

By engaging volunteers in data collection and analysis, citizen science helps advance extensive research efforts and fosters scientific literacy. However, these data must be validated for reliable use in research, conservation, and policy-making. Although the number of global studies evaluating the reliability of citizen science data is on the rise, few such studies have been conducted in Türkiye. One validation tool is Benford's Law. In this study, we applied Benford's Law alongside a Chi-Square Goodness-of-Fit Test to a dataset containing observation counts for 533 bird species reported in Türkiye, as cataloged on the "eBird Türkiye" [27] platform. Verifying data in this manner is crucial to ensuring accuracy and reliability, which in turn enables better decision-making for biodiversity protection in the framework of Conservation Biology.

2. Materials and Methods

2.1. Study Material

We employed a data set from the citizen science platform eBird Türkiye [27], containing observation counts for 533 bird species recorded in Türkiye. Established in 2002 by the Cornell Lab of Ornithology in the United States, eBird has become the largest citizen science platform focused on bird sightings worldwide. With more than 600 million bird observations, it maintains one of the most substantial biodiversity databases, creating extensive opportunities for large-scale research on migration routes, population changes, and the effects of environmental shifts [5].

2.2. Benford Law

Benford's Law was used to assess this data set's reliability. According to the first-digit test, Benford's Law states that the leading non-zero digit of a number—or a set of values representing a given phenomenon—is not uniformly distributed. A high percentage of numbers begin with 1, followed by 2, 3, and so forth. Specifically, 30.1% of data points start with 1, 17.6% with 2, 12.5% with 3, 9.7% with 4, 6.7% with 5, 6.4% with 6, 5.1% with 7, 4.9% with 8, and 4.7% with 9, indicating that smaller leading digits arise more frequently than larger ones [15, 16]. For the first-digit test, Benford's Law is applied using the formula:

$$P(d) = \log_{10} (1 + 1/d) \qquad (2.1)$$

Here it is:

P(d): The probability that the first digit of a number is the digit "d".

d: The digit representing the first digit of the number. This digit can be an integer from 1 to 9 (i.e. $d \in \{1,2,...,9\}$).

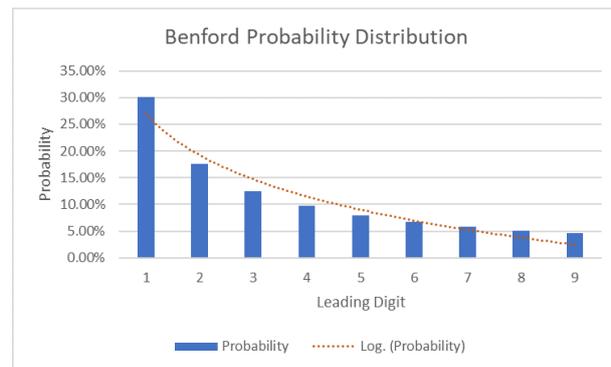$\log_{10}$: Denotes the logarithm to base 10.

Figure 2. Probability distributions of the first digits according to Benford's Law
Note: Prepared from [28]

2.3. Data Analysis

We completed the following steps in Microsoft Excel to verify the data's conformity with Benford's Law:

2.3.1. Forming the Data Set from the Database: A .txt file was exported from the eBird database (see Appendix 1). This file was imported into MS Excel, after which bird species and their observation counts were filtered to create the final data set.

2.3.2. Extracting the First Digit: The first digit of each observation count in the data set was identified.

2.3.3. Calculating Frequencies: The observed frequency of each leading digit was recorded.

2.3.4. Calculating the Percentage Frequency: The observed frequency was divided by the total data count to obtain the percentage frequency.

2.3.5. Computing Benford Percentages: The "=LOG$_{10}$(number)" formula was used to determine each digit's expected probability under Benford's Law.

2.3.6. Calculating Expected Frequencies According to Benford's Law: Expected frequencies were derived from the total number of observations in conjunction with each digit's Benford percentage.

2.3.7. Conducting the Chi-Square Goodness-of-Fit Test

: This test identifies whether the observed variable matches a given distribution. The formula used was:

$$x^2 = (Observed - Expected)^2 / Expected \qquad (2.2)$$

2.3.8. Calculating of the p-Value

The sum of the values found in the Chi-Square Test of Likelihood was found. Then the p value was found as "0.19865" by using the formula "=CHI-SQUARED.DISTR.RIGHT (x; degrees of freedom)". The degrees of freedom used in the formula is one less than the 9 variables used, i.e. the degrees of freedom is 8. The calculation was made as follows: =CHI-SQUARED.DISTR.RIGHT(11,05;8)

*If p-value < 0.05, the data do not follow Benford's Law.

*If the p-value is ≥ 0.05, the data obey Benford's Law.

According to these conditions, since the p-value is 0.19868 > 0.05, the data obey Benford's Law.

3. Results and Discussion

In this study, the conformity of e-bird observation data in Turkey with Benford's Law was examined. The results of the analysis showed that the first digits of the observation counts for each species were distributed in a manner similar to the frequency distribution predicted by Benford's Law.

3.1. Key Findings

3.1.1. Overall Fit: A strong relationship was found between the frequency percentages of the first digits of the observation numbers (from 1 to 9) and the percentages predicted by Benford's Law. In particular, it was found that small digits such as 1, 2, and 3 appeared with a higher frequency, as predicted by Benford's Law.

Table 1 shows that the most common digit value is the number 1 (%32,8) and the least common digit value is the number 9 (%2,8) This is a situation seen in the first digit test in Benford's Law.

3.1.2. Statistical Tests: Statistical tests (e.g., chi-square test) have shown that the difference between observed and expected frequencies is not statistically significant. This confirms that the data set is generally consistent with Benford's Law.

3.1.3. Differences: Small deviations from Benford's Law predictions have been observed for some numbers. However, it has been determined that these deviations are not significant enough to distort the overall structure of the data set.

The table shows that the absolute difference between the Benford Percentage and the percentage occurance of the digits (frequency percentage) varies between %2,7 – 0,1. This is another indication that the phenomenon complies with Benford's Law. Looking at the graph in Figure 3, the biggest difference between the Frequency Percentage and the Benford Percentage (2,7%) is seen in the number 1. The smallest difference between the Frequency Percentage and the Benford Percentage(0,1 %) is seen in the number 8.

In short, the Frequency Percentage shows how often the first hoseholds in your dataset actually occur, while the Benford Percentage expresses how often these households should appear in a natural distribution. The difference between these two percentage indicates how well your data set fits Benford's Law and whether there are potentially any anomalies. The small numbers seen in the difference show that our data seti is compliance with Benford's Law.

Figure 3 presents a bar chart of the observed frequency percentage (actual data) and the expected percentage under Benford's Law (predicted data); the close alignment between these bars illustrates consistency with Benford's Law.
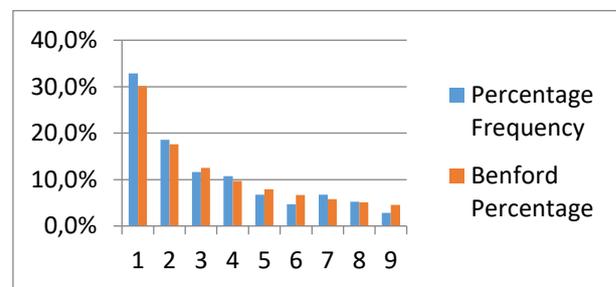

Figure 3. Conformity of Observation Counts for 533 Bird Species from eBird Türkiye to Benford's Law.

Table 1 summarizes the application of Benford's Law and the Chi-Square Goodness-of-Fit Test to the data set. As shown in Table 1, the Chi-Square Goodness-of-Fit Test produced a x^2 value of 11.05, with a p-value of 0.19865. Because this p-value is greater than 0.05, the data fit Benford's Law.

The findings show that e-bird observation data exhibit a distribution consistent with Benford's Law, which is characteristic of naturally occurring data sets. This result provides strong evidence that the data set has not been manipulated and was created as a result of a natural process.

M. Afsar

4. Conclusion and Discussion

The reason why Benford's Law is used in the naturalness calculation of ecosystems is that its message is mathematical, that is, it is clear.

The Benford probability curve in reality means that a whole army of numbers contains "process information". This "process information" does not explain what the numbers that go into the probability calculation are and what formulas are used to derive them. Here "process information" simply means numbers with a background.

This message of Benford's Law is as clear as its theoretical probability values [24].

What is meant here by numbers with a resume?

These numbers mean the system in other words. If the numbers that make up a whole conform to Benford's Law, there is a natural consciousness that allows that whole to exist and to continue to exist. In other words, if a whole has a natural consciousness, it is expected to obey Benford's Law. In this case, the term "system" is used instead of "whole".

Table 1. Summary of Steps and Results in MS Excel

Number	Frequency	Percentage Frequency	Benford Percentage	Expected Benford Frequency	Observed – Expected	(Observed – Expected)^2	Chi-Square Goodness-of-Fit Test
1	175	32,8%	30,1%	160,45	14,55	211,73	1,32
2	99	18,6%	17,6%	93,86	5,14	26,45	0,28
3	62	11,6%	12,5%	66,59	-4,59	21,09	0,32
4	57	10,7%	9,7%	51,65	5,35	28,59	0,55
5	36	6,8%	7,9%	42,20	-6,20	38,48	0,91
6	25	4,7%	6,7%	35,68	-10,68	114,12	3,20
7	36	6,8%	5,8%	30,91	5,09	25,91	0,84
8	28	5,3%	5,1%	27,26	0,74	0,54	0,02
9	15	2,8%	4,6%	24,39	-9,39	88,15	3,61
Toplam	**533,00**	**1,00**	**1,00**	**533,00**			**11,05**

This study examined observational data for 533 bird species from the "eBird Türkiye" citizen science platform using Benford's Law.

According to the Chi-Square Goodness-of-Fit Test, the data conform to Benford's Law, suggesting that the dataset has not been manipulated and can serve as a credible source for scientific research. However, while Benford's Law offers a meaningful overview of data consistency, it does not provide an absolute guarantee of data accuracy [16].

Given the ongoing concerns about data quality in citizen science applications, this finding is quite important. Although errors, omissions, or systematic biases may be present in data provided by volunteers, the results obtained in this study support the statistical reliability of such data. Similarly, various studies [26,29,30] have demonstrated that volunteer science data can be used reliably when evaluated using careful assessment methods.

The Benford's Law used in the study is a powerful and practical tool for analyzing natural data distributions, especially in data sets containing a large number of observations. This method has been successfully applied in the analysis of biological data (e.g., cell counts, taxonomic distributions, genome sizes) in the past. In this context, applying the Benford test to eBird Turkey data provides an important methodological contribution to

evaluating the quality of volunteer science data in our country.

However, some minor deviations were also observed during the analysis. In particular, differences were found between the observed and expected frequencies in some steps (e.g., numbers 1 and 9). Although these differences are not statistically significant, investigating the causes of such deviations in future studies may contribute to the development of improvement strategies to enhance data quality. For example, systematic biases such as the over-reporting of certain species or the re-entry of data by some observers could be responsible for these differences.

Another important contribution of the study is that a data verification method of this scale has been applied in Turkey. Although citizen science studies are becoming widespread around the world, studies testing the reliability of such data in Turkey are quite limited. This situation indicates that the study not only provides results regarding the reliability of eBird data but also has the potential to increase the scientific validity of citizen science applications in Turkey.

However, there are some limitations to this study. First, the analysis was conducted solely on the basis of the first-order distribution. For a more comprehensive evaluation of the data, second-order analysis, time series analysis, or spatial validation methods could also be used. Furthermore, it should be noted that Benford's Law alone is not sufficient for measuring the reliability of data, but it does provide an effective method for preliminary screening and general evaluation.

The number of digits of large ecological data such as the number of cells in cyanobacteria colonies[19] and pollen counts [20] are quite high. For this reason, while the second and third step test applications of Benford's Law are tested in data sets with large digit numbers of this type, it is predicted that the first step test will be sufficient considering the digit numbers of the number values in the data set in the study and the literatüre studies. The reason for applying the second and the third step test in Benford's Law is to detect errors and anomalies more precisely.

Our findings emphasize the need for more rigorous data validation measures to further enhance the accuracy and reliability of citizen science data. While Benford's Law serves as an effective preliminary screening tool, additional validation methods – such as expert verification, inter – observer agreement and outlier detection – may be employed to improve data quality [25,29].

In essence, citizen science democratizes scientific knowledge production while simultaneously supporting biodiversity monitoring and conservation activities. Expanding citizen science initiatives in Türkiye could strengthen the country's research base and broaden conservation participation among the public.

Citizen science—which engages members of the public in data collection and analysis—contributes significantly to large-scale research efforts and fosters scientific literacy. Data reliability is especially critical in fields such as Conservation Biology, where accurate information can directly inform conservation measures. As one of the few studies evaluating the trustworthiness of citizen science data in Türkiye, this paper sheds light.

Benford's Law is only an indicator. A deviation from Benford's Law is a warning sign that there may be anomalies or manipulation in a data set. However, this does not necessarily mean that the data is incorrect. It simply indicates that further examination and in-depth research is required. In short, while Benford's Law is a powerful audit and analysis tool, it has important limitations that must be carefully considered in terms of its application areas and the characteristics of the data set. If these limitations are ignored, misleading results may be obtained.

In conclusion, this study proposes an applicable and valid method for evaluating the quality of data obtained from volunteer science platforms in Turkey, while also supporting the usability of this data in scientific studies. Future studies can conduct more in-depth evaluations by analyzing data sets separated according to different species groups, observers, or time periods.

5. Recommendations

Strengthening Citizen Science in Türkiye: The continued development of citizen science projects, especially in natural and environmental sciences, could greatly enhance Türkiye's research capacities.

Validating Data in Other Citizen Science Platforms: Similar assessments should be undertaken on other platforms across Türkiye, since the data generated often serve as a primary resource for biodiversity monitoring.

Integrating Citizen Science into Education: Introducing citizen science projects in the educational system can promote scientific thinking among young

people and heighten overall awareness of environmental issues.

Unlocking the Potential of Citizen Science: Citizen science approaches can address various local challenges in Türkiye, such as water quality monitoring, urban land-use mapping, and seismic event tracking. Establishing a dynamic citizen science community may offer both societal and scientific advantages, paving the way for a more robust future.

Acknowledgment

Author's Contributions

Çetin ÇELİK: Draft and wrote the manuscript, analyzed the data and generated the result with the data.

Murat AFSAR: Draft and wrote the manuscript.

Ethics

There are no ethical issues after the publication of this manuscript.

References

[1]. İçen, H., Kapan, K., & Uysal, A. (2024). Bilimsel araştırma süreçlerinde yeni bir yaklaşım olarak vatandaş bilimi (citizen science). Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, 14(3), 1583-1597.

[2]. Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

[3]. Cooper, C. B., Shirk, J., & Zuckerberg, B. (2014). The invisible prevalence of citizen science in global research: migratory birds and climate change. PLoS ONE, 9(9), e106508. https://doi.org/10.1371/journal.pone.0106508.

[4]. Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., ... & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation, 181, 236-244.

[5]. Sullivan, B.L., C.L. Wood, M.J. Iliff, R.E. Bonney, D. Fink, and S. Kelling. 2009. eBird: a citizen-based bird observation network in the biological sciences. Biological Conservation 142: 2282-2292.

[6]. Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., ... & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. Biological conservation, 213, 280-294.

[7]. Pocock, M. J., Chandler, M., Bonney, R., Thornhill, I., Albin, A., August, T., ... & Danielsen, F. (2018). A vision for global biodiversity monitoring with citizen science. In Advances in ecological research (Vol. 59, pp. 169-223). Academic Press.

[8]. Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature communications, 6(1), 1-8.

[9]. Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., ... & Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. Biological Conservation, 208, 113-120.

[10]. Hortal, J., Jiménez-Valverde, A., Gómez, J. L. G., Lobo, J. M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos, 117(6), 847-858. https://doi.org/10.1111/j.0030-1299.2008.16434.x.

[11]. Szabo, J. K., Khwaja, N., Garnett, S. T., & Butchart, S. H. (2012). Global patterns and drivers of avian extinctions at the species and subspecies level.

[12]. Ward, D. F. (2014). Understanding sampling and taxonomic biases recorded by citizen scientists. Journal of insect conservation, 18, 753-756.

[13]. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. Scientific reports, 7(1), 9132.

[14]. Primack, R. B. (2006). Essentials of conservation biology (Vol. 23). Sunderland: Sinauer Associates.

[15]. Newcomb, S. (1881). Astronomical observatories. Science, (59), 377-380.

[16]. Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, 551-572.

[17]. Cerri, J., Testa, F., & Rizzi, F. (2018). The more I care, the less I will listen to you: How information, environmental concern and ethical production influence consumers' attitudes and the purchasing of sustainable products. Journal of Cleaner Production, 175, 343-353.

[18]. Perazzoni, F., Bacelar-Nicolau, P., & Painho, M. (2020). Correction: Perazzoni, F.; Bacelar-Nicolau, P.; Painho, M. Geointelligence against Illegal Deforestation and Timber Laundering in the Brazilian Amazon. ISPRS Int. J. Geo-Inf. 2020, 9, 398. ISPRS International Journal of Geo-Information, 9(10), 573.

[19]. Costas, R., & Bordons, M. (2008). Is g-index better than h-index? An exploratory study at the individual level. Scientometrics, 77(2), 267-288.

[20]. Docampo, S., del Mar Trigo, M., Aira, M. J., Cabezudo, B., & Flores-Moya, A. (2009). Benford's law applied to aerobiological data and its potential as a quality control tool. Aerobiologia, 25, 275-283.

[21]. Friar, J. L., Goldman, T., & Pérez–Mercader, J. (2012). Genome sizes and the Benford distribution. PLoS One, 7(5), e36624.

[22]. Campos, L., Salvo, A. E., & Flores-Moya, A. (2016). Natural taxonomic categories of angiosperms obey Benford's law, but artificial ones do not. Systematics and Biodiversity, 14(5), 431-440.

[23]. Özkundakci, D., & Pingram, M. A. (2019). Nature favours "one" as the leading digit in phytoplankton abundance data. Limnologica, 78, 125707.

[24]. Özkan K. Estimating ecosystem naturalness using Benford's Law and Generalized Benford's Law. Turkish Journal of Forestry. Haziran 2021;22(2):73-82. doi:10.18182/tjf.907217

[25]. Michener, W. K. (2006). Meta-information concepts for ecological data management. Ecological informatics, 1(1), 3-7.

[26]. Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., ... & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment, 10(6), 291-297.

[27]..eBird. (2024). eBird: An online database of bird distribution and abundance [https://ebird.org/data/download?p=ebd_TR_202301_202312_relOct-2024.zip]. eBird, Cornell Lab of Ornithology, Ithaca, New York. Available: http://www.ebird.org. (Accessed: October 25, 2024).

[28]. Türkyener, C.M. "Benford Yasası ve Mali Denetimde Kullanımı", Sayıştay Dergisi, Ocak-Mart 2007, 64:111-122.

[29]..Fournier, A. M., Sullivan, A. R., Bump, J. K., Perkins, M., Shieldcastle, M. C., & King, S. L. (2017). Combining citizen science species distribution models and stable isotopes reveals migratory connectivity in the secretive V irginia rail. Journal of Applied Ecology, 54(2), 618-627.

[30]. Kosmala, M., Wiggins, A., Swanson, A., and Simmons, B. (2016). Assessing data quality in citizen science. Front Ecol Environ 2016; 14(10): 551–560, doi:10.1002/fee.1436

M. Afsar

Appendices
Appendix 1: Data Set Used in the Study

Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı
Acanthis flammea	1	Anatidae sp. (dabbling duck sp.)	214	Asio otus	553	Cercotrichas galactotes	543	Columba sp.	22	Emberiza caesia	718	Gallinago media	29
Accipiter nisus	8050	Anatidae sp. (teal sp.)	16	Astur gentilis	236	Certhia brachydactyla	1231	Coracias garrulus	1893	Emberiza calandra	19324	Gallinago sp.	1
Accipitridae sp. (eagle sp.)	9	Anser albifrons	2688	Athene noctua	1767	Certhia familiaris	145	Corvus corax	12480	Emberiza cia	1070	Gallinula chloropus	11805
Accipitridae sp. (hawk sp.)	1125	Anser anser	3631	Aythya ferina	32751	Certhia familiaris/brachydactyla	16	Corvus cornix	75733	Emberiza cineracea	62	Garrulus glandarius	20648
Accipitriformes/Falconiformes sp.	24	Anser cygnoides	1	Aythya fuligula	10089	Ceryle rudis	1161	Corvus frugilegus	24791	Emberiza cirlus	4166	Gavia arctica	1448
Acridotheres tristis	2118	Anser erythropus	635	Aythya marila	6	Cettia cetti	7196	Corvus sp.	12	Emberiza citrinella	3878	Gavia stellata	15
Acrocephalus agricola	129	Anser serrirostris	11	Aythya nyroca	2485	Charadriiformes sp. (shorebird sp.)	45	Corvus sp. (crow sp.)	28	Emberiza citrinella x leucocephalos	4	Gelochelidon nilotica	2341
Acrocephalus arundinaceus	2351	Anser sp. (Domestic type)	16	Aythya sp.	400	Charadriidae sp.	16	Coturnix coturnix	389	Emberiza hortulana	1856	Geronticus eremita	1119
Acrocephalus melanopogon	604	Anthropoides virgo	145	Botaurus minutus	581	Charadrius hiaticula	5089	Crex crex	16	Emberiza leucocephalos	4	Glareola nordmanni	2300
Acrocephalus palustris	713	Anthus campestris	1707	Botaurus stellaris	187	Charadrius/Thinornis/Anarhynchus sp.	2	Cuculus canorus	1003	Emberiza melanocephala	1867	Glareola pratincola	3151
Acrocephalus palustris/scirpaceus	46	Anthus cervinus	1699	Branta leucopsis	43	Chlidonias hybrida	4825	Curruca cantillans	334	Emberiza pusilla	5	Glareola pratincola/nordmanni	2
Acrocephalus schoenobaenus	641	Anthus japonicus	72	Bubo bubo	37	Chlidonias leucopterus	6220	Curruca communis	3349	Emberiza rustica	12	Grus grus	11333
Acrocephalus scirpaceus	3468	Anthus pratensis	11995	Bucanetes githagineus	15	Chlidonias niger	563	Curruca conspicillata	16	Emberiza schoeniclus	4076	Gulosus aristotelis	6977
Acrocephalus scirpaceus/arundinaceus	2	Anthus pratensis/spinoletta	2	Bucanetes mongolicus	14	Chlidonias sp.	4	Curruca crassirostris	393	Emberiza sp.	87	Gymnoris xanthocollis	86
Acrocephalus sp.	222	Anthus richardi	22	Bucephala clangula	55	Chloris chloris	15515	Curruca curruca	15515	Eremophila alpestris	1176	Gypaetus barbatus	204
Actitis hypoleucos	4374	Anthus sp.	500	Burhinus oedicnemus	915	Chroicocephalus genei	7384	Curruca melanocephala	4234	Erithacus rubecula	20427	Gyps fulvus	850
Aegithalos caudatus	9154	Anthus spinoletta	4171	Buteo buteo	91910	Chroicocephalus genei/ridibundus	115	Curruca mystacea	364	Eudromias morinellus	163	Haematopus ostralegus	1159
Aegolius funereus	9	Anthus trivialis	2187	Buteo lagopus	7	Chroicocephalus ridibundus	185807	Curruca nisoria	872	Falco columbarius	289	Halcyon smyrnensis	598
Aegypius monachus	1309	Apodidae sp.	4	Buteo rufinus	3989	Chroicocephalus ridibundus/Ichthyaetus melanocephalu	71	Curruca ruppeli	497	Falco eleonorae	595	Haliaeetus albicilla	294
Aerospiza/Tachyspiza/Accipiter/Astur sp.	118	Apus affinis	225	Buteo sp.	64	Ciconia ciconia	560426	Cursorius cursor	53	Falco naumanni	2225	Hieraaetus pennatus	2766
Alaemon alaudipes	16	Apus apus	31512	Calandrella brachydactyla	3033	Ciconia nigra	15269	Cyanistes caeruleus	12328	Falco naumanni/tinnunculus	65	Himantopus himantopus	17257
Alauda arvensis	12120	Apus apus/pallidus	868	Calcarius lapponicus	52	Ciconia sp.	11	Cygnus atratus	1	Falco peregrinus	784	Hippolais icterina	42
Alaudala heinei	382	Apus pallidus	3914	Calidris alba	918	Circaetus gallicus	17565	Cygnus columbianus	14354	Falco sp.	114	Hippolais languida	222
Alaudala rufescens	391	Apus/Tachymarptis sp.	295	Calidris alpina	24860	Circus aeruginosus	10207	Cygnus columbianus/cygnus	28	Falco sp. (small falcon sp.)	5	Hippolais olivetorum	109
Alaudala rufescens/heinei	1	Aquila chrysaetos	460	Calidris canutus	522	Circus cyaneus	614	Cygnus cygnus	680	Falco subbuteo	1811	Hippolais sp.	15
Alaudidae sp.	63	Aquila fasciata	19	Calidris falcinellus	220	Circus macrourus	211	Cygnus olor	3060	Falco tinnunculus	6199	Hirundinidae sp.	707
Alcedo atthis	3764	Aquila heliaca	484	Calidris ferruginea	2611	Circus macrourus/pygargus	19	Cygnus sp.	10	Falco vespertinus	970	Hirundo rustica	121192
Alectoris chukar	2358	Aquila nipalensis	82	Calidris minuta	15342	Circus pygargus	721	Delichon urbicum	22147	Ficedula albicollis	1307	Hirundo rustica x Delichon urbicum	2
Alopochen aegyptiaca	4	Aquila sp.	20	Calidris pugnax	30337	Circus sp.	55	Dendrocopos leucotos	179	Ficedula hypoleuca	455	Hydrobates pelagicus	22
Amandava amandava	1	Ardea alba	11820	Calidris sp.	228	Cisticola juncidis	378	Dendrocopos major	651	Ficedula hypoleuca/albicollis/semitorquata	18	Hydrocoloeus minutus	2486
Ammomanes deserti	14	Ardea cinerea	35126	Calidris sp. (peep sp.)	507	Clamator glandarius	170	Dendrocopos major/syriacus	21	Ficedula parva	862	Hydroprogne caspia	811
Ammoperdix griseogularis	298	Ardea cinerea/purpurea	1	Calidris temminckii	210	Clanga clanga	364	Dendrocopos sp.	59	Ficedula semitorquata	146	Ichthyaetus audouinii	248
Anarhynchus alexandrinus	7109	Ardea ibis	6203	Calonectris diomedea	506	Clanga pomarina	194711	Dendrocopos syriacus	4980	Ficedula sp.	11	Ichthyaetus ichthyaetus	1875
Anarhynchus asiaticus	7	Ardea purpurea	3967	Caprimulgus aegyptius	1	Clanga pomarina/clanga	103	Dendrocoptes medius	579	Francolinus francolinus	718	Ichthyaetus melanocephalus	6259
Anarhynchus leschenaultii	144	Ardea sp.	7	Caprimulgus europaeus	428	Coccothraustes coccothraustes	1264	Dryobates minor	925	Fringilla coelebs	85856	Iduna caligata	3
Anas acuta	8851	Ardeidae sp.	11	Carduelis carduelis	48687	Coloeus monedula	110839	Dryocopus martius	81	Fringilla montifringilla	2469	Iduna pallida	3801
Anas crecca	172593	Ardeola ralloides	4580	Carpodacus erythrinus	897	Columba livia	105231	Egretta garzetta	25515	Fringillidae sp.	28	Irania gutturalis	411
Anas platyrhynchos	71682	Arenaria interpres	437	Carpospiza brachydactyla	104	Columba oenas	783	Egretta/Ardea sp.	46	Fulica atra	443372	Jynx torquilla	398
Anatidae (duck sp.)	1968	Argya altirostris	458	Cecropis rufula	9280	Columba palumbus	23238	Elanus caeruleus	169	Galerida cristata	22237	Ketupa zeylonensis	1
Anatidae sp.	40	Asio flammeus	42					Emberiza buchanani	30	Gallinago gallinago	5720	Laniidae sp.	63

Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı	Tür Adı	Gözlem Sayısı
Lanius collurio	12817	Melopsittacus undulatus	1	Oenanthe sp.	43	Phylloscopus humei	70	Pterocles orientalis	1013	Sterna hirundo	5958	Turdus sp.	101
Lanius excubitor	97	Mergellus albellus	14	Oenanthe xanthoprymna	24	Phylloscopus inornatus	4	Ptyonoprogne rupestris	3678	Sterna sp.	1	Turdus torquatus	343
Lanius isabellinus	14	Mergus merganser	18	Oriolus oriolus	1364	Phylloscopus nitidus	201	Puffinus yelkouan	88982	Sterninae sp.	100	Turdus viscivorus	3780
Lanius minor	1258	Mergus serrator	1235	Otis tarda	231	Phylloscopus orientalis	283	Pycnonotus leucotis	71	Sternula albifrons	2535	Tyto alba	248
Lanius nubicus	1285	Merops apiaster	51541	Otus brucei	26	Phylloscopus proregulus	26	Pycnonotus xanthopygos	7346	Streptopelia decaocto	34292	Upupa epops	5617
Lanius senator	2014	Merops nubicoides	4	Otus scops	641	Phylloscopus sibilatrix	374	Pyrrhocorax graculus	2334	Streptopelia orientalis	1	Vanellus gregarius	21
Larinae sp.	1435	Merops persicus	1752	Oxyura leucocephala	6135	Phylloscopus sindianus	1131	Pyrrhocorax pyrrhocorax	4871	Streptopelia sp.	12	Vanellus indicus	177
Larus argentatus	48	Microcarbo pygmaeus	19761	Pandion haliaetus	367	Phylloscopus sp.	72	Pyrrhula pyrrhula	1149	Streptopelia turtur	4718	Vanellus leucurus	17
Larus armenicus	23660	Milvus migrans	12707	Panurus biarmicus	2356	Phylloscopus trochilus	3784	Rallus aquaticus	1752	Strigiformes sp.	11	Vanellus spinosus	6586
Larus cachinnans	1716	Milvus milvus	44	Paridae sp.	2	Phylloscopus trochilus/collybita	56	Recurvirostra avosetta	7717	Strix aluco	373	Vanellus vanellus	29376
Larus cachinnans/michahellis	102	Milvus milvus/migrans	2	Parus major	28122	Pica pica	44830	Regulus ignicapilla	1096	Sturnus vulgaris	391248	Xenus cinereus	15
Larus cachinnans/michahellis/armenicus	1695	Monticola saxatilis	340	Passer domesticus	118523	Picidae sp.	11	Regulus regulus	1993	Sylvia atricapilla	8003	Zapornia parva	227
Larus canus	463	Monticola solitarius	1170	Passer domesticus/hispanio	649	Picus canus	121	Regulus regulus/ignicapilla	7	Sylvia borin	1038	Zapornia pusilla	13
Larus fuscus	1378	Montifringilla nivalis	1156	Passer domesticus/montanu	24	Picus viridis	1000	Remiz pendulinus	2653	Sylvia/Curruca sp.	77		
Larus marinus	39	Morus bassanus	5	Passer hispaniolensis	31250	Platalea leucorodia	7426	Rhodopechys sanguineus	334	Tachybaptus ruficollis	27473		
Larus michahellis	186373	Motacilla alba	32703	Passer moabiticus	486	Plectrophenax nivalis	38	Rhodospiza obsoleta	1252	Tachymarptis melba	42792		
Larus sp.	2898	Motacilla cinerea	3104	Passer montanus	15613	Plegadis falcinellus	19263	Riparia riparia	29302	Tachyspiza badia	8		
Limosa lapponica	99	Motacilla citreola	266	Passer sp.	71	Pluvialis apricaria	2127	Rissa tridactyla	57	Tachyspiza brevipes	8142		
Limosa limosa	8908	Motacilla flava	28620	Passeriformes sp.	1844	Pluvialis fulva	17	Saxicola maurus	1050	Tachyspiza brevipes/Accipiter nisus	3		
Limosa sp.	7	Motacilla flava/tschutschensis	2	Pastor roseus	17377	Pluvialis squatarola	7615	Saxicola rubetra	3582	Tadorna ferruginea	91380		
Linaria cannabina	14223	Motacilla sp.	190	Pelecanus crispus	4965	Podiceps auritus	78	Saxicola rubicola	4498	Tadorna tadorna	15947		
Linaria flavirostris	430	Muscicapa sp.	8	Pelecanus onocrotalus	40275	Podiceps cristatus	37633	Saxicola rubicola/maurus	26	Tetraogallus caspius	141		
Locustella fluviatilis	22	Muscicapa striata	4890	Pelecanus sp.	60	Podiceps grisegena	91	Scolopax rusticola	79	Tetrax tetrax	3		
Locustella luscinioides	341	Neophron percnopterus	1000	Perdix perdix	266	Podiceps nigricollis	4957	Serinus pusillus	737	Thalasseus bengalensis	45		
Locustella naevia	31	Netta rufina	10096	Periparus ater	5268	Podiceps sp.	10	Serinus serinus	6331	Thalasseus sandvicensis	8673		
Loxia curvirostra	573	Numenius arquata	3884	Pernis apivorus	12912	Podicipedidae sp.	8	Sitta europaea	1322	Thinornis dubius	4204		
Lullula arborea	4579	Numenius phaeopus	293	Pernis ptilorhynchus	1	Poecile lugubris	812	Sitta krueperi	1366	Tichodroma muraria	40		
Luscinia luscinia	485	Numenius phaeopus/arquata	2	Petronia petronia	5883	Poecile palustris	299	Sitta neumayer	2840	Tringa erythropus	3019		
Luscinia luscinia/megarhynchos	16	Numida meleagris	1	Phalacrocoracidae sp.	165	Porphyrio poliocephalus	749	Sitta sp.	4	Tringa glareola	8316		
Luscinia megarhynchos	3464	Nycticorax nycticorax	7115	Phalacrocorax carbo	113833	Porphyrio sp. (swamphen sp.)	30	Sitta tephronota	258	Tringa nebularia	5427		
Luscinia svecica	2216	Oena capensis	24	Phalacrocorax carbo/Gulosu	336	Porzana porzana	49	Somateria mollissima	2	Tringa ochropus	5113		
Lymnocryptes minimus	82	Oenanthe cypriaca	80	Phalaropus fulicarius/lobat	2	Prinia lepida	2145	Spatula clypeata	24849	Tringa sp.	19		
Lyrurus mlokosiewiczi	87	Oenanthe deserti	29	Phalaropus lobatus	212	Prunella collaris	74	Spatula querquedula	30509	Tringa stagnatilis	1410		
Mareca penelope	28129	Oenanthe finschii	943	Phasianus colchicus	540	Prunella modularis	1151	Spatula querquedula/Anas cre	150	Tringa totanus	10826		
Mareca strepera	3459	Oenanthe isabellina	3372	Phoeniconaias minor	1	Prunella ocularis	114	Spilopelia senegalensis	10817	Troglodytes troglodytes	5528		
Marmaronetta angustirostris	15	Oenanthe lugens	2	Phoenicopterus roseus	359938	Psittacula eupatria	3263	Spinus spinus	19325	Turdus atrogularis	3		
Melanitta fusca	164	Oenanthe melanoleuca	2585	Phoenicurus ochruros	6358	Psittacula eupatria/krameri	16	Stercorarius parasiticus	131	Turdus iliacus	1603		
Melanitta nigra	87	Oenanthe monacha	57	Phoenicurus phoenicurus	4167	Psittacula krameri	8672	Stercorarius parasiticus/poman	9	Turdus merula	32522		
Melanocorypha bimaculata	150	Oenanthe oenanthe	6707	Phoenicurus sp.	1	Psittacula sp.	714	Stercorarius pomarinus	13	Turdus philomelos	11318		
Melanocorypha calandra	14059	Oenanthe pleschanka	7	Phylloscopus collybita	17489	Pterocles alchata	2	Stercorarius sp. (jaeger sp.)	1	Turdus pilaris	5089		